Nested CTR–PTR Modeling for Sparse Purchase Prediction in Sponsored Search

Greg Kocher

eBay, USA

AdKDD Workshop at KDD 2026 • August 10, 2026 • Jeju, Korea

Why purchase prediction is hard

CPA sponsored search needs an accurate, **well-calibrated** $p(\text{sale} \mid x)$ — the purchase-through-rate (PTR).

- **Data sparsity:** only a tiny fraction of impressions lead to a sale; click labels are denser but weaker. Probabilities are often 10^{-3}–10^{-2} or smaller.
- **Sample selection bias (SSB):** the classical factorization

$$\underbrace{p(z=1 \mid x)}_{p_{\text{PTR}}} = \underbrace{p(y=1 \mid x)}_{p_{\text{CTR}}} \times \underbrace{p(z=1 \mid y=1, x)}_{p_{\text{CVR}}}$$

  trains post-click CVR on a *clicked-only* subset, but serving scores the *entire* impression population $\Rightarrow$ inference-time mismatch.

Prior approaches: entire-space multi-task models

- **ESMM**: jointly models CTR, CVR, and their product over the full impression space.
- **ESM^2, AITM**: densify supervision via intermediate post-click actions or sequential dependence across conversion tasks.
- **ESCM^2, DCMT**: identify biases of the product architecture itself (IEB, PIP) and fix them with counterfactual debiasing.

Powerful, but architecture-dependent: custom multi-task heads, hybrid losses with per-task weights, counterfactual machinery. Hard to apply when CTR and PTR models are built by different pipelines, teams, or model families.

Goal: much of the same benefit but without as much added complexity.

The rare-event trick: $\operatorname{logit}(p) \approx \ln p$

Taking logs of the factorization is **exact**:

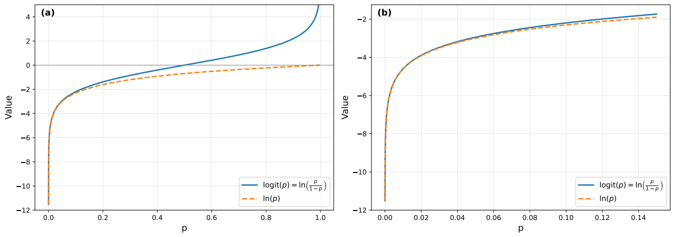
$$\ln p_{\mathrm{PTR}}(x) = \ln p_{\mathrm{CTR}}(x) + \ln p_{\mathrm{CVR}}(x)$$

For small p:
$\operatorname{logit}(p) = \ln p - \ln(1 - p) \approx \ln p$

$$\operatorname{logit}(p_{\mathrm{PTR}}(x)) \approx \underbrace{s_{\mathrm{CTR}}(x)}_{\text{CTR prior}} + \underbrace{g(x)}_{\text{learned residual}}$$

The **multiplicative** factorization becomes **additive in logit space**.

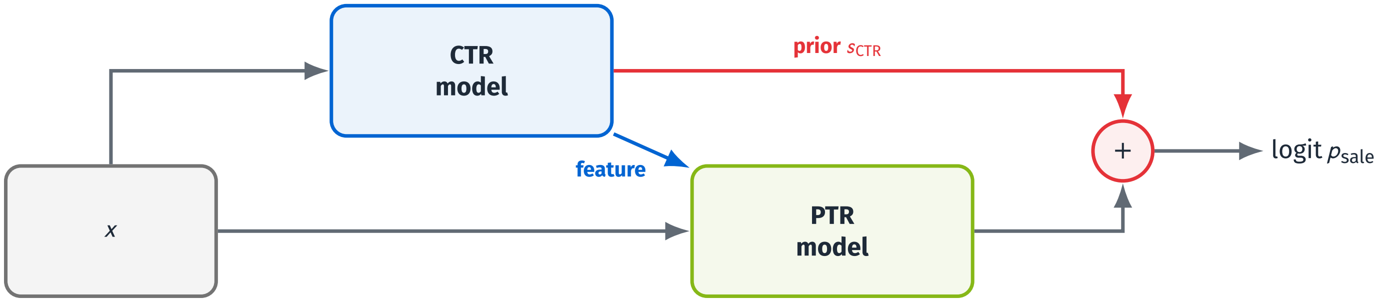
Closed-form error:
$\operatorname{logit}(p) - \ln p = -\ln(1 - p) = p + \frac{p^2}{2} + O(p^3)$



p	$\operatorname{logit}(p)$	$\ln(p)$	% Diff.
0.001	−6.907	−6.908	0.014
0.010	−4.595	−4.605	0.219
0.050	−2.944	−2.996	1.742
0.100	−2.197	−2.303	4.795

Nested CTR–PTR: inject the click signal

An upstream CTR model is trained on the **entire impression space**; its logit score $s_{\mathrm{CTR}}(x)$ feeds the downstream PTR model two ways:



- **Model-agnostic**: works for any models that produce / consume a logit (LR, GBDT, neural nets, etc.).
- Both models see **all impressions** — no clicked-only CVR head.

Variant 1 — CTR as a prior

The CTR score is a **fixed logit-space offset**; the sale model learns a residual correction:

$$\underbrace{\operatorname{logit}\big(p(z{=}1 \mid x)\big)}_{\text{sale logit}} = \underbrace{s_{\mathrm{CTR}}(x)}_{\text{CTR prior}} + \underbrace{g_{\mathrm{PTR}}(x)}_{\text{learned correction}}$$

- Factorization-like interpretation in logit space: CTR gives the initial estimate, PTR learns the click $\rightarrow$ sale adjustment — conceptually similar to a post-click conversion model.
- Logit-offset priors are already standard practice, e.g. for position bias.
- **Latency-friendly**: CTR and PTR can be evaluated **in parallel** and summed.

Variants 2 & 3 — CTR as a feature (and both)

2. CTR as a feature — arbitrary nonlinear use of the click signal:

$$\underbrace{\operatorname{logit}\big(p(z{=}1 \mid x)\big)}_{\text{sale logit}} = f_{\mathrm{PTR}}\big(x,\ \underbrace{s_{\mathrm{CTR}}(x)}_{\text{CTR feature}}\big)$$

3. CTR as prior + feature — click-informed initialization *and* learned interactions:

$$\underbrace{\operatorname{logit}\big(p(z{=}1 \mid x)\big)}_{\text{sale logit}} = \underbrace{s_{\mathrm{CTR}}(x)}_{\text{CTR prior}} + h_{\mathrm{PTR}}\big(x,\ \underbrace{s_{\mathrm{CTR}}(x)}_{\text{CTR feature}}\big)$$

- The feature pathway can distinguish *Exploratory Clicks* (eye-catching, rarely bought) from high-intent clicks — and sidesteps ESMM's PIP problem.

Why this design

- **Entire-space training** for both models — avoids the clicked-only SSB without a constrained CVR head.
- **Model-agnostic & modular:** CTR and PTR keep their own data, features, pipelines, and teams.
- **No hybrid loss:** multi-task signal without tuning per-task weights in a combined objective.
- **Nonlinear click–sale interactions** instead of a rigid multiplicative decomposition.
- **Calibrated for the sale task** — the model directly targets $p(\text{sale} \mid x)$ over the full impression space.

Offline results

Held-out test set: >170M impressions, >10M clicks, >400K sales — relative change vs. the single-task $p(\text{sale})$ baseline (nested variants shaded):

Model	ROC-AUC	PR-AUC	RIG	LogLoss	Cal. Err.
P(sale)–Baseline	0.00%	0.00%	0.00%	0.00%	0.00%
P(sale)–Large	+0.09%	+0.91%	+0.58%	−0.11%	+26.06%
Multiplicative (CTR×CVR)	+0.13%	−0.66%	+0.55%	−0.11%	−58.51%
MTL–Joint	+0.16%	+0.82%	+0.88%	−0.16%	+84.57%
CTR Prior Only	+0.27%	+1.57%	+1.54%	−0.30%	−11.17%
CTR Feature Only	+0.26%	+2.38%	+1.62%	−0.31%	−8.51%
CTR Prior + Feature	+0.29%	+2.28%	+1.80%	−0.34%	−43.62%

- Gains are consistent across metrics and hold vs. the **equal-capacity** P(sale)–Large baseline.
- Multiplicative CTR×CVR improves ROC-AUC but *degrades* PR-AUC; nested variants avoid this.

Online A/B tests & launch

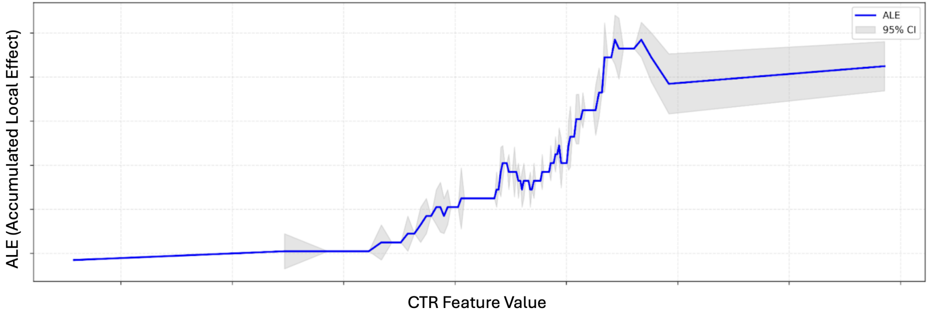
Several weeks of live session-randomized eBay traffic, relative to the production ranking policy:

Model	Revenue Metric 1	User Metric 1	Platform Metric 1
CTR Prior + Feature	+0.29% (p=.20)	**+0.33% (p=.04)**	**+0.48% (p=.03)**

- Statistically significant gains in user metrics tied to purchase rates and platform metrics tied to sales volume; directionally positive ad revenue.

A Nested CTR-PTR variant was launched to production at eBay.

Interpretability



Accumulated Local Effects (ALE) of the CTR feature inside the PTR model (*CTR as a Feature*).

- With CTR injected, post-click-conversion features become most important.
- ALE: the CTR feature's effect on the PTR score rises nearly monotonically, then plateaus and dips at high CTR — consistent with *Exploratory Clicks*: eye-catching items that attract **clicks but not purchases**.
- Caveat: wide confidence intervals in the sparse high-CTR region — suggestive, not conclusive.

Takeaways

- A simple approximation, $\text{logit}(p) \approx \ln p$, turns CTR×CVR into an **additive logit-space decomposition** — no custom multi-task architecture needed.
- Inject the dense upstream click signal as a **prior**, a **feature**, or both — model-agnostic, modular, and effective.
- Gains **offline**, **online**, and deployed at eBay.

Greg Kocher • eBay